International Journal of Human-Computer Studies 144 (2020) 102500

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Sonification of the self vs. sonification of the other: Differences in the sonification of performed vs. observed simple hand movements

Roberto Bresin*[a], Maurizio Mancini[b], Ludvig Elblaus[a], Emma Frid[a]

[a] KTH Royal Institute of Technology, Stockholm, Sweden
[b] University College Cork, Cork, Ireland

ARTICLE INFO

Keywords:
Sonification
Mapping
Hand movement
Performance
User role

ABSTRACT

Existing works on interactive sonification of movements, i.e., the translation of human movement qualities from the physical to the auditory domain, usually adopt a predetermined approach: the way in which movement features modulate the characteristics of sound is fixed. In our work we want to go one step further and demonstrate that the user role can influence the tuning of the mapping between movement cues and sound parameters. Here, we aim to verify if and how the mapping changes when the user is either the performer or the observer of a series of body movements (tracing a square or an infinite shape with the hand in the air). We asked participants to tune movement sonification while they were directly performing the sonified movement vs. while watching another person performing the movement and listening to its sonification. Results show that the tuning of the sonification chosen by participants is influenced by three variables: role of the user (performer vs observer), movement quality (the amount of Smoothness and Directness in the movement), and physical parameters of the movements (velocity and acceleration). Performers focused more on the quality of their movement, while observers focused more on the sonic rendering, making it more expressive and more connected to low-level physical features.

1. Introduction

Movement sonification is the translation of human movement qualities (e.g., amplitude, speed) from the physical to the auditory domain (see Dubus and Bresin, 2013 for an overview). Such a translation can be achieved, for example, through a direct mapping between movement and audio features. An “optimal” mapping will effectively communicate the same meaning across performers (i.e., humans performing dance movements) and listeners (i.e., humans listening to the resulting translation in the auditory domain). The goal of the present paper is to provide insights in the design of effective movement-to-sound mappings. In particular, we hypothesize that the interactive sonification mapping (Hunt and Hermann, 2011) designed by a person who is performing the sonified movement will be different, in general, from the one designed by a person who is just listening to the sonification. In other words, we aim to answer to the following question: in interactive sonification, will the mapping between movement and sound be different if it is designed by the person actually performing the sonified movement vs. being designed by the person who is observing/listening the same sonified movement?

We know from previous research on the analysis of gestures in music performance that body and instrument constrains can influence both expressive performances and musicians’ body movements (Dahl et al., 2009), and that we can consider music performers as experts in translating gestures into sounds (Godøy et al., 2005). In a previous work on the execution of *legato* piano performances, Repp (1995, 1997) demonstrated that participants perform differently when playing the piano themselves (performing condition) versus controlling the note duration of the same pre-recorded note sequence (listening condition). In both conditions, participants were asked to perform an optimal minimal and maximal *legato*. In piano performance *legato* articulation is commonly achieved by keeping the keys corresponding to two consecutive *legato* notes pressed for a short time interval called “Key-Overlap Time”(KOT). Repp found that performers used larger KOTs than listeners, and that KOTs tend to be negative (corresponding to the absence of physical overlapping of the piano keys) for minimal optimal legato and for keys corresponding to low frequencies, since their corresponding tones had a longer decay time perceptually corresponding to an acoustic *legato*. A possible explanation is that when controlling the pre-recorded sequence, participants cognitively focused on the acoustic feedback only, while when performing the legato pianists focused more on the physical action that produces both the

* Corresponding author
E-mail addresses: roberto@kth.se (R. Bresin), m.mancini@cs.ucc.ie (M. Mancini), elblaus@kth.se (L. Elblaus), emmafrid@kth.se (E. Frid).

https://doi.org/10.1016/j.ijhcs.2020.102500
Received 16 June 2019; Received in revised form 16 June 2020; Accepted 18 June 2020
Available online 20 June 2020

mechanical overlap of piano keys and the acoustic overlap of their corresponding sounds.

Other previous works on the expression and recognition of emotions in both speech and music showed that the Brunswikian Lens Model (Brunswick, 1956) has been "used in several fields to study how observers correctly and incorrectly use objective cues to perceive physical or social reality" (Juslin and Timmers, 2010; Scherer, 1978). In particular, in music performance, performers and listeners can use cues from the same set of acoustical, haptic, and visual cues in different ways when communicating and decoding expression, respectively (Juslin and Timmers, 2010). In a previous study, we found a small but significant effect of movement-to-sound mapping on movement qualities (energy, smoothness, directness), that changed depending on the sound model (Frid et al., 2016).

In this paper we aim to test whether or not a sonification model is tuned differently by participants driving the sonification with their own movement, compared to participants who sonify the movement performed by another person they are observing. In addition, we want to investigate if the tuning of this sonification depends also on the quality of the movements. In particular, we aim to confirm (or deny) the following research hypotheses:

H1 - Effect of condition on sound parameters: people tune sound parameters of an interactive movement sonification model differently, in terms of the use of one or more parameters, when performing a movement themselves vs. when they are observing another person performing a movement. More in detail, we expect that the observer will adjust the sound parameters for highlighting some of the characteristics of the observed movements, while the sonification made by the performer will be more homogeneous.

H2 - Effect of movement on sound parameters: The type of movement (e.g., a smooth curved trajectory vs. a jerky straight one) influences the way people tune sound parameters. In particular, we expect smooth movements to be sonified with more dull and low frequent sounds, while jerky trajectories will be portrayed with sounds characterized by higher frequencies and louder sound level.

In other words, we hypothesize that, like in Repp's studies (Repp, 1995; 1997), participants will focus more on physical actions when performing the movement themselves, compared to when they only observe the performed movements.

2. Experiment

2.1. Experimental design

The experimental setting used in our study is illustrated in Fig. 1. Participants were randomly assigned to two groups: the Performance Group (PG) and Observation Group (OG). Each group was assigned to one of two adjoining rooms, separated by a one-way mirror, as illustrated in Fig. 1. Participants in the OG group were assigned to the Observation Room (OR) and could observe participants in the PG group, who were assigned to the Performance Room (PR). PG participants in the PR room could not observe OG participants in the OR room. Both rooms had an identical setup consisting of a computer with headphones and a multi-touch tablet with sliders.

2.1.1. Performance room

PG participants performed their task in the PR room. The room was equipped with a motion capture system, and participants were fitted with a set of motion capture markers that they were holding with their dominant hand (see Fig. 2).

The equipment used in this room consisted of:

- 1 computer displaying instructions and showing the two different hand movements to be performed by participants in the PR room. The two movements represented the shapes of a square and an infinity symbol, respectively[1]).
- 1 server computer for running the motion capture software system (Arena software[2]), logging the motion capture data, and streaming movement low-level features to the client computers in the PR and OR in real-time.
- 1 client computer running software for real-time sound synthesis (implemented with SuperCollider[3]), connected via USB to a tablet running a four-slider interface (implemented with TouchOSC[4]) used for manipulating the sound synthesis and logging the slider data incoming from the tablet. The client software also controlled the GUI on the tablets (e.g., activating specific sliders), provided functions for proceeding to the next stage of the experiment (this was manually controlled by the experiment supervisor, see Section 2.3), and implemented the routine for pacing the experiment (e.g., of trials), as controlled by the experimenter.
- 1 pair of Beyer Dynamic DT770 Pro headphones, worn by PG participants.
- 1 video camera that was used for recording each PG participant.

2.1.2. Observation room

The room was equipped with a one-way mirror that enabled OG participants to observe the movements performed by PG participants in the PR room. The equipment used in this room was:

- 1 computer displaying the same instructions as in the PR room.
- 1 computer with identical client setup as the computer in the PR room.
- 1 pair of identical headphones as in the PR room.
- 1 video camera that was used for recording each OG participant.

2.2. Participants

A total of 25 people took part in the experiment (14F, 11M; Mean = 22.88 years, SD = 2.89). Participants were students of the Degree Programme in Media Technology at KTH Royal Institute of Technology. The experiment was performed during three days. The experiment lasted 25 min, on average.

2.3. Procedure

The 25 participants were divided into two groups: 14 PG participants (7F, 7M), and 11 OG participants (7F, 4M), for a total of 11 pairs, and 3 single participants in the PR (2M, 1F). For each observer/performer pair of participants, the entrance to the respective room took place at separate times.

Each PG participant, whose arm movements had to be sonified, was holding a wand with her/his dominant hand. The wand comprised a handle similar to that of a screw driver and included reflective markers on its top (see Fig. 2). We chose not to have markers placed directly on the hand of participants in order to force them to produce very similar movements between each participant and consequently to have a reduced number of degrees of freedom in the movements. The markers were tracked in 3D space with an OptiTrack motion capture system by Natural Point.[5] Three low-level movement features were computed in realtime: (1) velocity of the user's hand, obtained by differentiating positional data; (2) acceleration, obtained by differentiating velocity; (3) jerk, obtained by differentiating acceleration. Low-level movement features were used in the sonification process presented in next

[1] The video can be found here: https://kth.box.com/v/DANCE2018

[2] https://optitrack.com/support/software/arena.html

[3] https://supercollider.github.io

[4] https://hexler.net/software/touchosc

[5] http://optitrack.com/products/flex-3

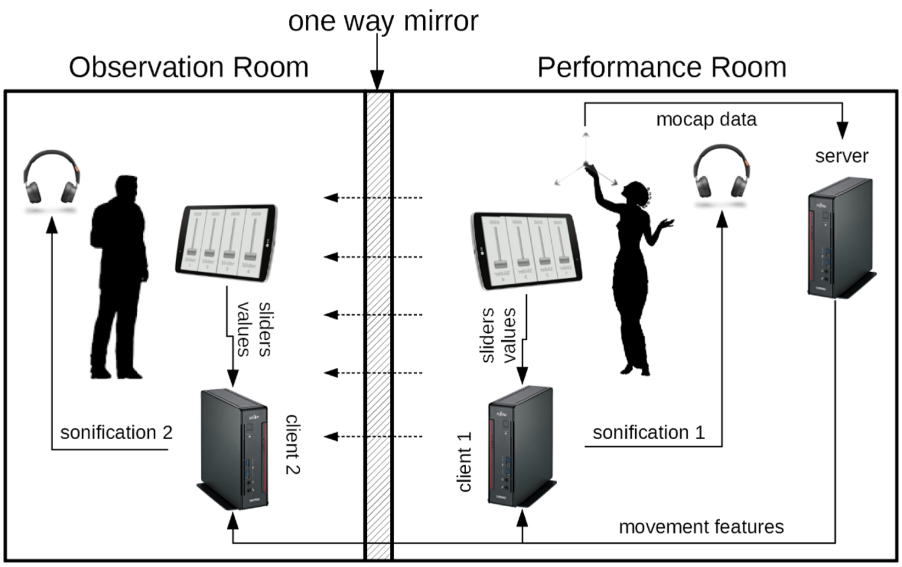

Fig. 1. A schematization of the experimental setting. Participants (performer and observer) are standing in two rooms separated by a one way mirror, so that the performer can be seen by the observer, while the opposite is not possible. A server computer receives the performer's motion capture data (i.e., the dominant hand position in space) and computes movement features, such as speed and acceleration, which are streamed to two client computers located in the two rooms. Both the performer and the observer can move four sliders displayed on a tablet. The sliders drive the movement sonification generated by the two client computers in each room.

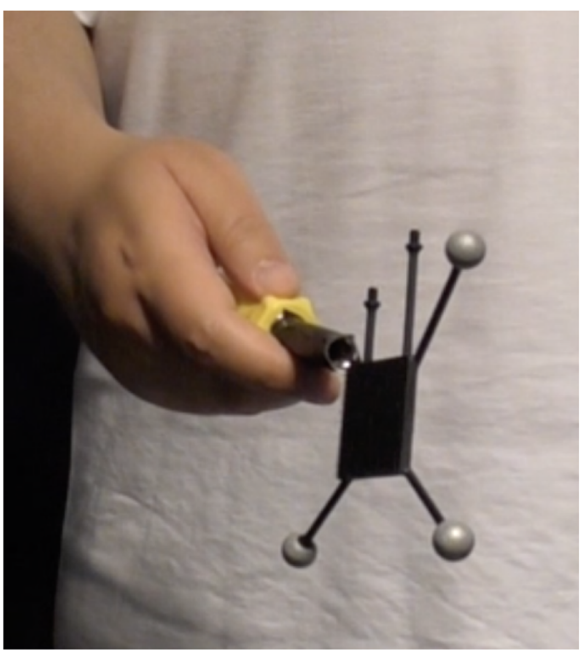

Fig. 2. The wand held by each PG participant with her/his dominant hand. The wand consisted of a handle similar to that of a screw driver, provided with reflective markers on its top.

sections.

Both PG and OG participants were presented with a set of four digital sliders on a multi-touch tablet computer (an Apple iPAD running TouchOSC), where each slider represented a weighting parameter for mapping a low-level movement feature to an audio feature, as presented in Table 1 (according to mappings reported in Dubus and Bresin (2013)) and implemented as explained in Section 2.4 (see Fig. 3 for a screenshot of the interface). For each participant, the starting position of each slider was automatically set to the middle of the scale, which corresponds to a 0 weight, i.e., no mapping. If a slider was set to its top position, i.e., its weight was set to 1, a full positive mapping would be used, that is, a high value in the movement's low-level feature will create a high value in the audio feature, according to the most commonly used polarizations used in previous studies (as reported in Dubus and Bresin (2013)). If the slider was set to its bottom position, the weight was set to -1, meaning that a full negative mapping would be used, i.e., the opposite of the most common polarizations. Between these endpoints, a fractional value representing a weight between -1 and 1 was associated with the position of the slider.

Participants received instructions by email one week prior to participating in the experiment, so that they could familiarize with the experiment procedure (full texts of the instructions are available in the annexes).

PG participants were instructed to perform the same sequence of movements as in a video presented on a monitor in front of them by moving approximately at the same speed as the person in the video. While doing their movements they had to adjust the slider(s) on a iPad so that the sound was affected by their movement in a way that they felt to be appropriate. They stopped when they were satisfied with the obtained sounds.

OG participants were instructed to observe through one-way mirror the movements performed by the PG participant in the other room. They were asked to adjust the slider(s) on an iPad until they felt that the sound affected by the movements of the PG participant was appropriate.

On the day of the experiment participants were asked to read the instructions again and fill out a consent form. Participants could ask questions to the instructors, if they had any. Once both participants were ready, the order of the sliders was randomly assigned to a synthesis parameter and the experiment started. It was organized according to the following six stages:

Table 1
The four movement features associated with the corresponding manipulated audio features and related mappings (as described in Dubus and Bresin, 2013).

Movement feature	Audio feature	Mapping type
Velocity of the user's dominant hand	Pitch	Low/high velocity to low/high pitch
Velocity of the user's dominant hand	Timbre and spectral slope	Low/high velocity to dark/bright timbre
Acceleration of the user's dominant hand	Periodicity of the sound	Low/high acceleration to periodic/aperiodic sound
Acceleration of the user's dominant hand	Amplitude	Low/high acceleration to low/high amplitude

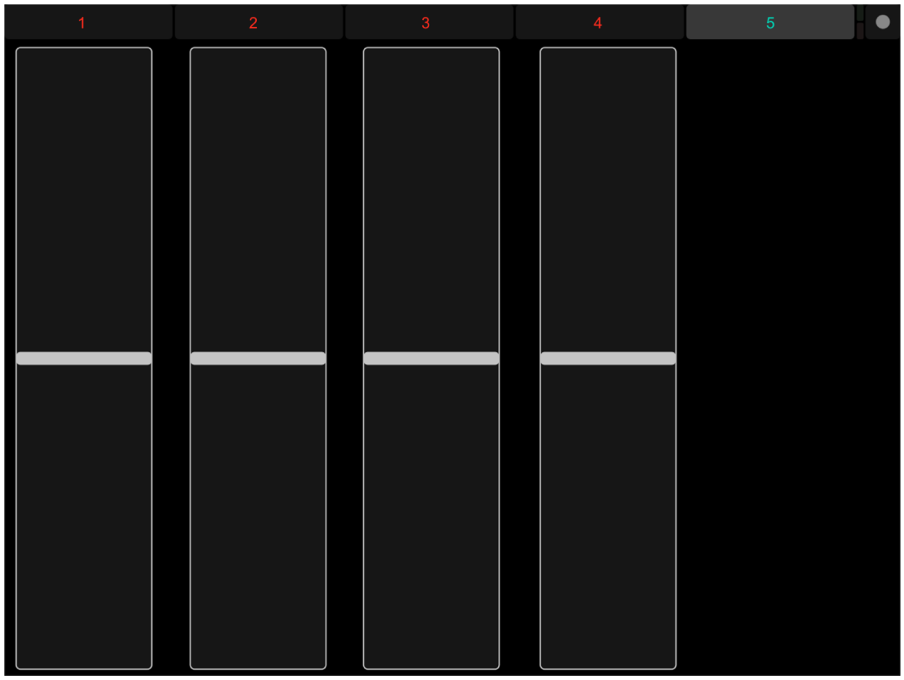

Fig. 3. Screenshot of the slider interface that was running on the tablet computers. The interface was the same for both PG and OG participants. The user interface was designed using black and grey not to distract participants with colours.

- *Stages 1–4:* Participants in both PG and OG groups controlled one slider at a time (Slider 1 during stage 1, Slider 2 during stage 2, and so on). At the beginning of each stage, all sliders were reset to the middle position. When both participants indicated that they were satisfied with the result, the experiment continued to the next phase (this was iterated for all four sliders).
  If participant in the PR room was satisfied with the sound before the participant in the OR room, the PR participant had to continue moving until OR participant was done.
- *Stage 5:* Both participants could modify all sliders at the same time. As for stages 1–4, they indicated when they were satisfied with the result.
- *Stage 6*: PR participant **only** was asked to repeat the movements until the instructor asked her/him to stop (approximately three iterations of the movement sequences). This was done to record audio of each participant's final sounds.

At the end of the experiment, both PG and OG participants exited the experiment rooms and were asked to fill out an online questionnaire about their participation.

The questionnaire was realized with the Survey Gizmo platform[6]. For the complete list of questions asked to participants please see the attached Annex Questionnaire.

2.4. Mapping movement to sound features

Participants in both groups could interactively tune the sonification of the movements produced by PG participants in the PR room by moving the sliders on the tablet.

The code used for generating the sounds for both PG add OG participants was written in SuperCollider.[7]

In the following sections we present the mapping between low-level movement features and sound features associated with each of the four sliders. The value of each slider was normalized between -1 and 1, where 0 corresponded to neutral behavior of the sound models.

2.4.1. Slider 1 - s1

Slider $s1$ mapped the velocity of movements to the fundamental frequency (F0) used in the sound models, according to the following rule:

- if the velocity of movement is equal to 0, then $F0 = 600\ Hz$;
- if the velocity of movement is greater than 0, then $F0$ is in the range $[20, 20000]\ Hz$ depending on the value of the slider.

The Super Collider code that mapped the value of $s1$ to $F0$ was:

```
f0 = (600 + (s1 * velocity * 600)).clip(20, 20000)
```

2.4.2. Slider 2 - s2

Slider $s2$ mapped the velocity of movement to the timbre used in the sound model, according to the following definition (written in Super Collider):

```
timbre = (0.5 + (s2 * 0.5 * velocity)) * 2.0
```

The resulting waveform used for the sound synthesis will be one of the following:

- *sinusoid*, if $timbre = 0$;
- *triangle*, if $timbre = 1$;
- *sawtooth*, if $timbre = 2$;

according to the following Super Collider code considering $s2$ value, and selecting the sound model `sig`:

```
sig = SelectX.ar(timbre, [SinOsc.ar(f0), LPF.ar
(LFTri.ar(f0), 11250), Saw.ar(f0)])
```

For values of timbre between 0 and 1, the resulting waveform will be a sinusoid cross-faded to a triangle waveform, and for values between 1 and 2, the waveform will be a triangle cross-faded to a

[6] https://www.surveygizmo.com

[7] The code is available here https://kth.box.com/v/DANCE2018

Table 2
Statistics (mean, median, standard deviation, 95% confidence lower and upper bound) of the two movement features Smoothness Index (SI) and Directness Index (DI) computed on the infinite and square movements separately (all PG participants).

	μ	$\tilde{x}$	σ^2	*lower*	*upper*
SI-square	0.05	0.00	0.02	0.02	0.08
DI-square	0.63	0.65	0.03	0.49	0.61
SI-infinite	0.29	0.18	0.09	0.13	0.31
DI-infinite	0.36	0.34	0.04	0.29	0.43

sawtooth.

2.4.3. Slider 3 - s3

The value of slider $s3$ mapped the acceleration of movements to the variations of the fundamental frequency ($F0$) determined by the position of $s1$. Low values of acceleration corresponded to small variations of $F0$, while high values of acceleration corresponded to large variations of $F0$. Frequency variations were produced by adding random values to $F0$ depending on the acceleration, as defined by the following Super Collider code:

```
f0 = f0 + LPF.ar(LFDNoise3.ar(f0, f0 * s3 * accelera-
tion), (f0 * 8).clip(800, 16000))
```

2.4.4. Slider 4 - s4

The value of slider $s4$ mapped the acceleration of movements to the variations of the sound level of the waveform determined by the position of $s2$. The resulting sound level is updated depending on the value of $s4$ and of the acceleration, according to the following Super Collider code:

```
sig = (sig * (1 -s4)) + (sig * s4 * acceleration * 4)
```

As a result, when $s4$ is left in its default position ($s4 = 0$), the sound level of the waveform is unvaried, while values greater than or less then 0 correspond to higher or lower sound level when acceleration is respectively larger or smaller.

All motion capture recordings and movements of the sliders for all participants were logged and saved for further analysis.

3. Analysis and results

Due to technical reasons, the data of 5 participants had to be disregarded. As a consequence, the total number of participants used in the analysis presented in the following sections was 20 (9F, 11M; Mean = 23 years, SD = 3.15), corresponding to 11 PG (4F, 7M) versus 9 OG participants (5F, 4M).

3.1. Movement features

In order to validate the movements performed by participants and also to investigate whether or not there was a correlation between movement characteristics and the sonification mappings that were set up by participants, we extracted 2 movement features on PG participants' dominant hand trajectory: the *Smoothness Index* (SI) and the *Directness Index* (DI). These features belong to layers 2 (SI) and 3 (DI) of the movement analysis framework described in Camurri et al. (2016). The framework consists of four layers, ranging from physical low-level signals (layer 1), to body joint movement (layer 2), to gesture (layer 3) and high-level qualities (layer 4). A detailed description of the framework is out of the scope of this paper. The two movement features have been chosen since they are related to the movement (SI) and gesture shape (DI) (Camurri et al., 2016). This is in line with the instructions provided to participants, as they were asked to focus on the shape and pace of two sample trajectories that were shown to them in a video before starting the experiment (see Section 2.3).

SI indicates how much a body joint (e.g., the user's dominant hand) is moving according to the specific laws of bio-mechanics defining smoothness, see Mazzarino and Mancini (2009), Hogan and Sternad (2007). SI has been described in the experiment in Frid et al. (2016), showing that it can be successfully exploited to sonify the trajectories of a group of children moving in a space monitored by a motion capture system. DI describes the shape of the trajectory of a body joint in reaching a target position in terms of directness vs. flexibility (Volpe and Camurri, 2011). Together with other movement mid-level features, it has been exploited in Ghisio et al. (2015) to create serious games using interactive sonification of movement in the rehabilitation of children with motor and cognitive impairment.

3.2. Movement features extraction

Movement features have been extracted on the movements performed by PG participants during stage 5 of the experiment, in which they performed the two sample movements (i.e., the infinite and the square) and adjusted all the four mapping sliders to the position resulting in the desired final sonic feedback. For each one of 11 PG participants, we manually segmented infinite and square movements. A segment corresponded to a sequence of consecutive movements of the same type (infinite or square), labeled by its type. Then, we automatically extracted SI and DI on the resulting 144 infinite and 113 square segments.

We computed the mean, median, standard deviation, variance and confidence intervals of SI and DI on all the segments. Table 2 reports these values for infinite and square segments separately. Then, we performed a paired two-tailed t test by comparing the values of SI and DI in square and infinite movements. The test showed significant differences ($p < 0.001$) between square and infinite movements, in terms of both SI and DI. As reported in Table 2, SI is higher in infinite movements, while DI is higher in square movements. This is true across all participants (see also Figs. 4, 5 and 6), with a confidence level of 95% and a low standard deviation.

Fig. 7 illustrates two examples of infinite (above) and square (below) segments, belonging to PG participants 3 and 8. The trajectories in the figure visually demonstrate the consistence between consecutive movements performed by participants.

Sound examples relative to the movements plotted in Figure 7 for both PG add OG participants can be found at the following link: https://kth.box.com/v/DANCE2018.

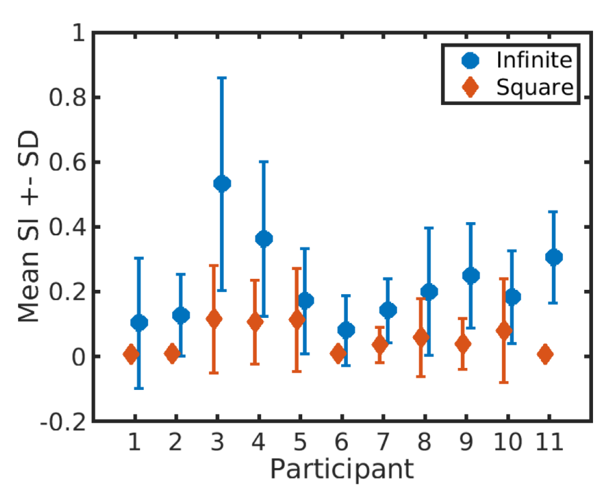

Fig. 4. Mean Smoothness Index (SI) computed per participant performing either the infinite or the square movement.

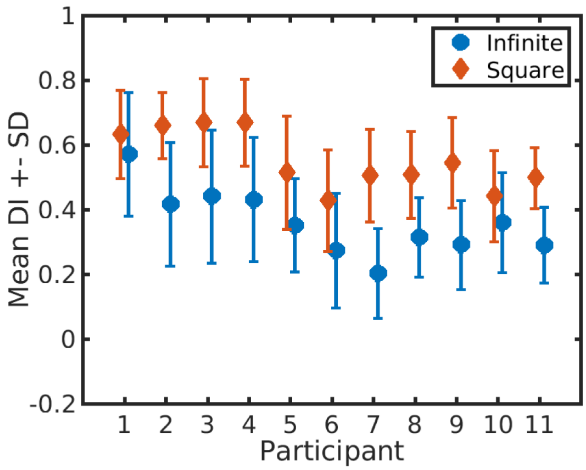

Fig. 5. Mean Directness Index (DI) computed per participant performing either the infinite or the square movement.

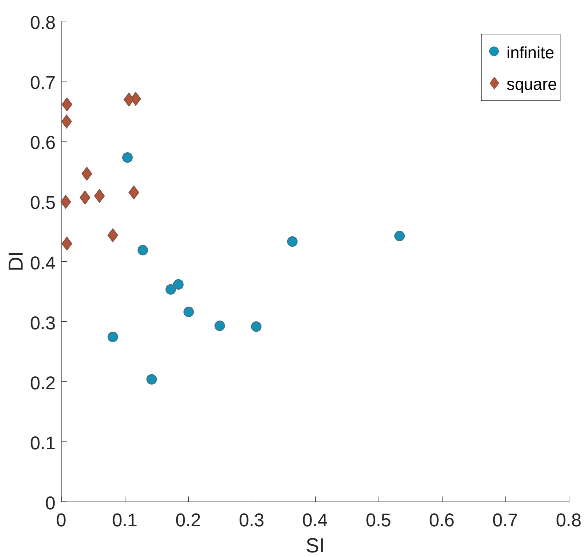

Fig. 6. Mean Smoothness Index (SI) and Directness Index (DI) computed on participants performing either the infinite (circle) or the square (diamond) movement. The graph clearly shows that we can identify two movement clusters in terms of SI and DI mean values.

3.3. Manipulation of sound model control parameters

The values assigned to each of the four sliders for the two groups of participants were analyzed to verify if participants made different choices in their sound settings (hypothesis H1), and if the sounds settings were correlated with specific mid-level movement features (hypothesis H2).

Each slider indirectly represented the question "How much (if any) influence do you think movement feature *x* should have on audio feature *y*?" which can provide an answer to hypothesis H1. The audio features are archetypal high-level characteristics that could be applied to a large number of sound models, making the results generalizable.

In Fig. 8 the medians, 25 and 75 percentiles for each slider are presented for the two groups, PG and OG respectively.

In order to verify the existence of different strategies for the control of sonification by the two groups of participants, we ran Pearson correlation tests between the final values of the sliders and the mean and median of the motion quality parameters for each group.

Performer Group (PG): Participants who made the sonification of own movements

C1 Values of slider $s1$ showed a significant negative correlation with the mean of SI of the infinite movement ($r(11) = -.625$, 95% CI, $p < 0.05$).
C2 Slider $s1$ showed a significant negative correlation with the variance of DI of the square movement ($r(11) = -.603$, 95% CI, $p < 0.05$).
C3 Slider $s2$ showed a significant positive correlation with slider $s3$ ($r(11) = 0.616$, 95% CI, $p < 0.05$).
C4 Slider $s4$ showed a significant positive correlation with the variance of SI of the infinite movement ($r(11) = 0.629$, 95% CI, $p < 0.05$).

Observer Group (OG): Participants who made the sonification observing the movement made by another participant

C5 Slider $s1$ showed a significant negative correlation with slider $s4$ ($r(9) = -0.723$, 95% CI, $p < 0.05$).
C6 Slider $s1$ showed a significant negative correlation with the mean velocity of the square movement ($r(9) = -0.678$, 95% CI, $p < 0.05$).
C7 Slider $s4$ showed a significant positive correlation with the standard deviation of DI of the square movement ($r(9) = 0.678$, 95% CI, $p < 0.05$).

3.4. Characterization of participants' behaviour

We wanted to verify the possible existence of clusters of participants with similar behaviour within the PG and OG groups. We did this by performing a two-step cluster analysis with participant groups (PG and OG) as categorical variable and the values of the four sliders as continuous variables. Two clusters characterize PG participants and are described mainly by values of sliders $s3$ and $s2$ (the two predictors with highest importance, 1.0 and 0.84 respectively). The first cluster is formed by eight participants and the other one by three. Three clusters of the same size (3 participants each) characterize OG participants, and are explained mainly by the values of sliders $s1$ and $s3$ (the two predictors with highest importance, 1.0 and 0.78 respectively). Results are shown in Fig. 9.

The results seem to suggest that both PG and OG participants can be grouped in two and three groups respectively. We are aware that the sample size is not large enough for generalizing the results of the cluster analysis, nevertheless the clusters that emerge show that the majority of PG participants had the same strategy when sonifying own gestures, while OG participants adopted three distinct tactics.

3.5. Strategy used by participants for solving the experimental task

In this section we summarize the main results from the online survey answered by both PG and OG participants after completing the experiment. The survey contained nine questions, but the most relevant for the current study was the one in which we asked participants to describe the strategy they had followed for solving the assigned task of fitting a sound to performed or observed movements.

Among PG participants, several participants (8 out of 14) explained their strategy by using terms related to their own body movements and how they were connected to sound. They focused more on the settings of the sliders and how these affected the sound resulting from their own movements. See Table 3 for a list of comments by eight PG participants.

Among OG participants, six out of eleven participants explained their strategy by using terms related to the movements made by the PG participants and how they were connected to sound. Five participants used terms related to emotions (e.g. feeling, pleasing, irritating). See Table 4 for a list of comments by eight OG participants.

4. Discussion

In this section, we discuss the results presented in the previous section. We first map the results to the two initial hypotheses and then

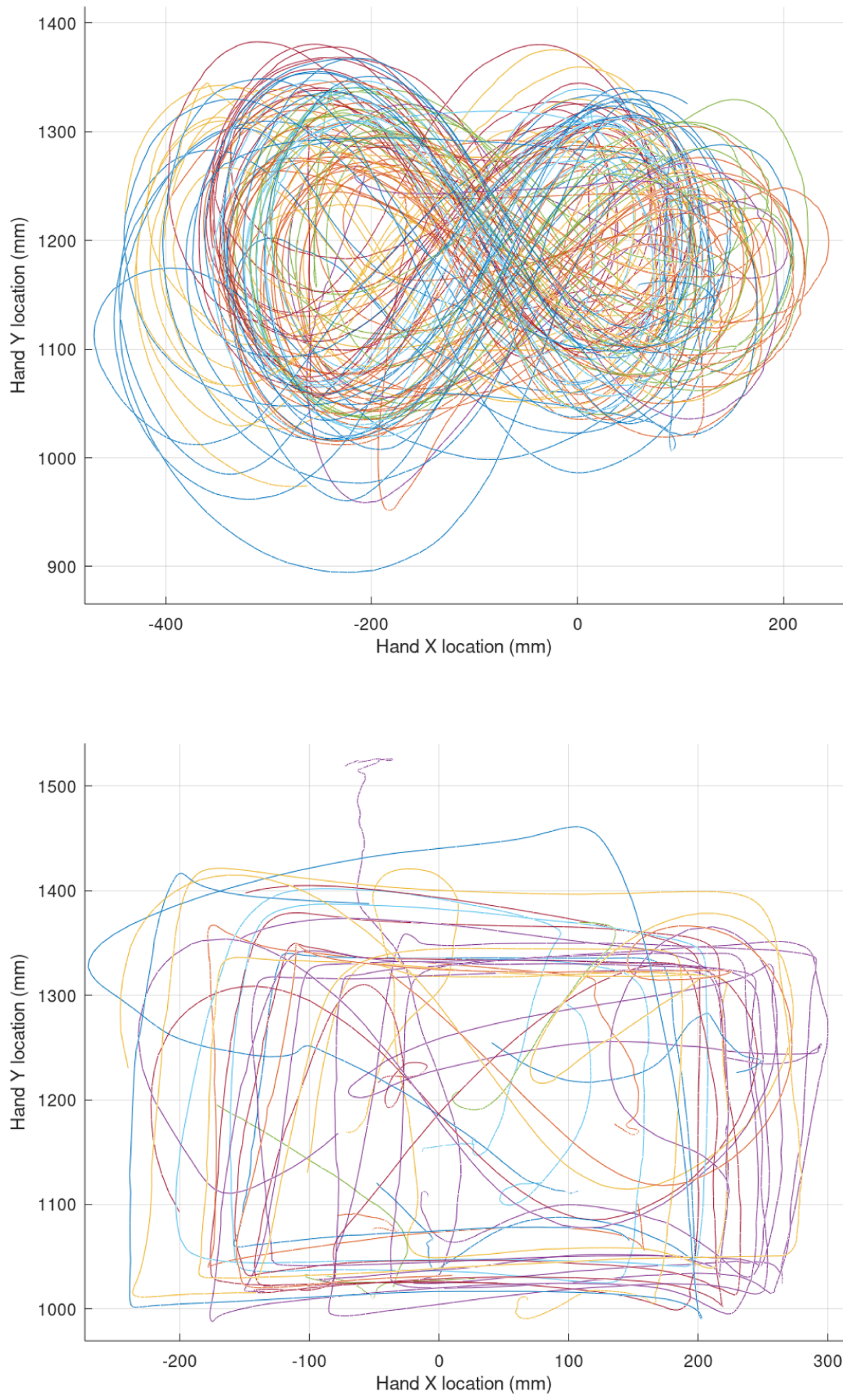

Fig. 7. PG participants 3 (above) and 8 (below), who took part in the experiment during the first day, performing the infinite and square movement. The two plots show the segments we provided as input to the movement feature extraction algorithms. Each segment has a different color.

we draw a more general interpretation of the results. In order to facilitate the reading of the discussion we have numbered the significant correlations presented in Section 3.3 from C1 to C7.

Hypothesis H1 - Effect of condition on sound parameters. From the results presented above it appears that there was an effect of condition on the manipulation of sound parameters by participants. It is reasonable to assume, by looking at Tables 3 and 4, that when sonifying hand movements, participants performing a gesture paid more attention to the information provided by own movements than participants observing the movements. Contrary to PG participants, OG participants made use of emotional terms for explaining their sonification strategy. This also seems to be confirmed by the cluster analysis presented in Fig. 9, in which PG participants seem to follow two very similar strategies, while OG participants adopted three different sonification strategies.

The analysis of the slider values presented in Section 3.3 also confirms that PG and OG participants used different strategies when adjusting the four sliders for generating the desired sonification. More in detail, PG participants focused more on sonifying the quality of their movements (smoothness and directness, see C1, C2 and C4), while OG participants focused more on low-level physical features of the movements, such as velocity and acceleration (see C5 and C6).

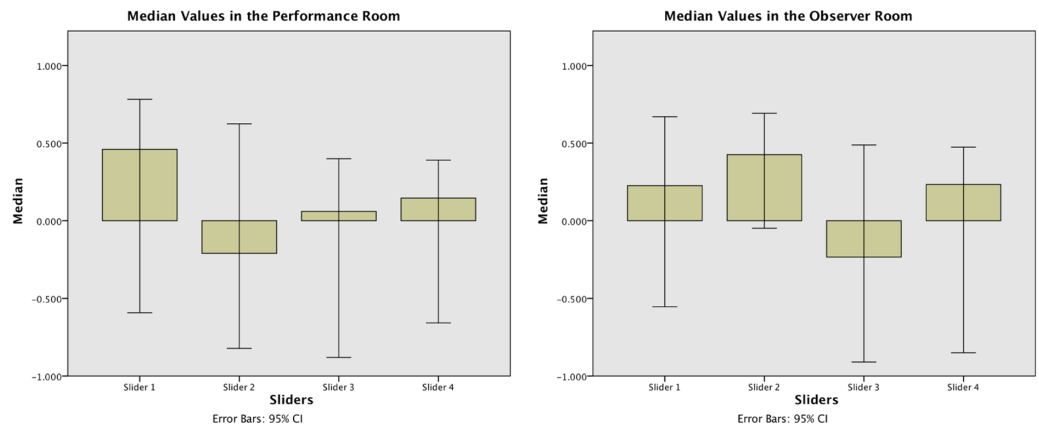

Fig. 8. Medians, 25 and 75 percentiles, and 95% confidence interval for each slider for PG participants (left panel) and OG participants (right panel).

Hypothesis H2 - Effect of movement on sound parameters. The analysis of the slider values presented in Section 3.3 shows differences between the two groups of participants. PG participants focused more on associating sound parameters (such as the sound level controlled by *s*4) to qualities of their own movement (i.e., variation of smoothness and directness). From the experiment, it emerges a tendency of PG participants towards a preference for sonifying the smoothness of own movements by using low frequency sounds, i.e., high values of smoothness were mapped to lower frequencies, below 600 Hz, and vice-versa (see C1). PG participants associated higher acceleration of movements to larger variations of sound level, which in turn were significantly correlated with larger variations of SI of the infinite movement (see C3). So, the faster the acceleration of the infinite movements, the louder their sonification. This is in line with principles of ecological perception and cross-modal correspondences (Spence, 2011).

OG participants associated higher acceleration of movements to larger variations of sound level, which in turn were significantly positively correlated with variations of standard deviation of the directness of the square movement (see C7). However, the most interesting difference between OG and PG participants' use of the sound parameters was that OG participants associated higher *F*0 with slower velocity and softer sound level (see C5 and C6), in contrast with the principles of ecological perception and cross-modal correspondences (Spence, 2011). This could be explained as a strategy by OG participants for emphasizing slower movements characterized by lower frequent sounds (low F0) with higher sound level in order to make them more perceivable (see C5 and C6).

These results confirm that our initial Hypotheses 1 was met and that

Table 3
Strategy used by eight PG participants for solving the experimental task.

Strategy
I used big moves to find out what the sliders did then I did more exact changes to get the sound I wanted.
I tried to find the position of the slider that I thought were the most fitting for both movements.
I compared my movements with the sound and tried to find the levels that was most likely my movements.
I tried to be consistent in my movements. I first increased the values of the sliders, whereafter I reduced the values, and then I allowed myself to try the places I liked. I tried to first do one slider at a time in the last step.
Trying to find the right volume and create a sound that reacted to my motions.
I tried to make the sound sounds harmonic, pacing my motion, and not too sharp to, somehow, explain how I was moving.
I tried to make the sound sync with my movements without delays.
I tried to see patterns with moving the sliders and then moving the stick, that was kind of difficult, but that was my strategy at least.

Hypothesis 2 was met by PG participants but not by OG participants. The latter used sonification strategies that are in conflict with principles of cross-modal correspondences and ecological perception, as documented in scientific literature (Dubus and Bresin, 2013; Spence, 2011).

Results also show a strong similarity with the findings reported in the work by Repp (1995, 1997) on the control of legato notes in piano performance. PG participants, like the pianists playing legato notes on a grand piano, focused more on the quality of their movement (smoothness and directness for PG participants, and optimal legato technique for pianists), while OG participants, like the pianists remotely controlling legato notes, focused more on the sonic rendering (making it more expressive and optimal for their ears and more connected to low-level

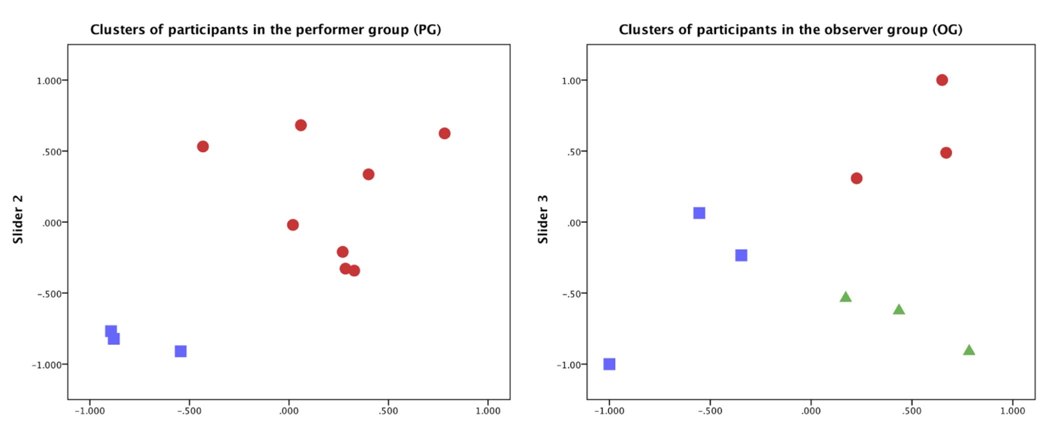

Fig. 9. Clusters of participants of the PG group (left panel) and the OG group (right panel) organized according to the values of two sliders.

Table 4
Strategy used by eight OG participants for solving the experimental task. The emotion column indicates strategies which involve emotional terms, while the motion column refers to strategies in which motion terms have been used by participants.

Strategy	Emotion	Motion
Based my answers on feeling.	x	
I just tested and went for whatever felt somewhat OK.	x	
I wanted to hear a difference when the rod was moved in the patterns, but also so it sounded as pleasing as possible.	x	x
I enjoyed hearing how she moved but I didn't like the sound so much. Nr 4 was the most irritating sound and nr 3 was the best way to hear her movements but still not a nice sound. I tried to find a balance where I got the "information" I wanted but not get it too uncomfortable for my ears.	x	x
Tried to adjust the sound so that the movement was palpable in the sound.	x	x
I listened for what setting made the most noise changes during movement in the first part. In the second part it was a bit easier to set the sliders in a way that seemed to fit with the actual movement.		x
I tried to find the timing of the movement and when the sensor stopped and turned direction, I looked for an clear pitch that differentiated from the constant noise.		x
I tried different settings until the sound was affected by the motion of the other person, in some way.		x

physical features, such as velocity, acceleration, or acoustic overlap as in the case of the Repp study).

It should be also noted that participants in the two groups made different choices, in average. In particular, sliders $s2$ and $s3$ were positioned in opposite locations of the scale (see Fig. 8). This confirms different strategies between the two groups; PG participants preferred a smoother timbre, characterized by a more stable fundamental frequency, compared to the OG participants. Indeed, a different use of the parameters associated with the sliders is confirmed by the different significant correlations between slider values and motion quality parameters as reported in the previous section (C1 to C7).

In summary, PG participants payed more attention to the smoothness and directness of their own movements when sonifying them, while OG participants focused more on velocity and acceleration of the movements they were watching. Also, the majority of PG participants used similar sonification strategies for their own gestures, while OG participants adopted three distinct tactics, as emerged from the cluster analysis on the use of the sliders.

The number of participants in our experiment is not large enough for the generalization of results. Nevertheless, the outcomes of our study suggest that designers of sonification of body movements should consider to use different strategies when sonifying observed vs. performed movements, in order to better support users' role.

5. Conclusions

Movement sonification consists in the translation of the physical properties of human movement (e.g., position, speed, acceleration of limbs) into audio parameters (e.g., volume, timbre, etc.). In the work presented in this paper, we investigated whether or not such a translation depends on the role of the individuals involved in the communication: they can be performers/listeners or observers/listeners. To do that, we asked two groups of people (performers vs. observers) to design a movement sonification of two types of movements: smooth (the infinite shape) vs. jerky (the square shape).

Results provide some preliminary support for our hypotheses: depending on their role, performers and observers will choose different sonification strategies. This is in line with previous findings described in Repp (1995, 1997), Frid et al. (2016). Also, the type of movement performed by the users (smooth/infinite vs. jerky/square) has an influence on the design of sonification depending on who was in control of the sonification.

A possible explanation of these results is that the sonification of own movements can be considered functional for supporting the movement itself, while the sonification of another person's movements focuses mainly on the communication of certain movement qualities.

In conclusion, results from our experiment show that two usually overlooked variables, that is, the role and context of users and the type of movements they perform, should be carefully addressed in the movement sonification design process. In this direction, further investigation with a larger number of participants could be carried out to verify these preliminary results and test them also in other contexts and with different sound models. Interesting scenarios could include, for example, the sonification of movements in sport and rehabilitation, or AR and VR applications, in which the representation of own body could be either embodied or disembodied.

CRediT authorship contribution statement

Roberto Bresin: Writing - original draft, Writing - review & editing, Data curation, Investigation, Conceptualization, Formal analysis, Methodology, Visualization, Supervision. **Maurizio Mancini:** Writing - original draft, Writing - review & editing, Data curation, Investigation, Software, Formal analysis, Methodology, Visualization. **Ludvig Elblaus:** Visualization, Investigation, Data curation, Software, Conceptualization, Methodology. **Emma Frid:** Writing - original draft, Data curation, Investigation, Conceptualization, Methodology.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property. We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from the Elsevier Editorial System.

Acknowledgements

This research was funded by European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 645553 ICT DANCE (IA, 2015–2017). Dr Maurizio Mancini carried out part of the movement analysis work described in the paper while being affiliated to DIBRIS, University of Genoa (Italy). The authors would like to thank the DANCE project coordinator Prof Antonio Camurri.

References

Brunswick, E., 1956. Perception and the Representative Design of Psychological Experiments. Berkeley: University

Camurri, A., Volpe, G., Piana, S., Mancini, M., Niewiadomski, R., Ferrari, N., Canepa, C., 2016. The dancer in the eye: towards a multi-layered computational framework of qualities in movement. Proceedings of the 3rd International Symposium on Movement and Computing. ACM, pp. 6.

Dahl, S., Bevilacqua, F., Bresin, R., Clayton, M., Leante, L., Poggi, I., Rasamimanana, N., 2009. Gestures in performance. In: Godøy, R.I., Leman, M. (Eds.), Musical Gestures: Sound, Movement, and Meaning. Routledge, New York, pp. 36–68.

Dubus, G., Bresin, R., 2013. A systematic review of mapping strategies for the sonification of physical quantities. PLoS ONE 8 (12), e82491.

Frid, E., Bresin, R., Alborno, P., Elblaus, L., 2016. Interactive sonification of spontaneous movement of children: cross-modal mapping and the perception of body movement qualities through sound. Front. Neurosci. 10. https://doi.org/10.3389/fnins.2016.00521.

Ghisio, S., Coletta, P., Piana, S., Alborno, P., Volpe, G., Camurri, A., Primavera, L., Ferrari, C., Guenza, C.M., Moretti, P., et al., 2015. An open platform for full body interactive sonification exergames. 2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN). IEEE, pp. 168–175.

Godøy, R.I., Haga, E., Jensenius, A.R., 2005. Playing “air instruments”: mimicry of sound-producing gestures by novices and experts. In: Gibet, S., Courty, N., Kamp, J. (Eds.), Gesture in Human-Computer Interaction and Simulation. Springer, Heidelberg, Germany, pp. 256–267.

Hogan, N., Sternad, D., 2007. On rhythmic and discrete movements: reflections, definitions and implications for motor control. Exp. Brain Res. 181 (1), 13–30.

Hunt, A., Hermann, T., 2011. Interactive sonification. In: Hermann, T., Hunt, A., Neuhoff, J.G. (Eds.), The Sonification Handbook. Logos Verlag, Berlin, Germany, pp. 273298.

Juslin, P.N., Timmers, R., 2010. Expression and communication of emotion in music performance. Handbook of Music and Emotion: Theory, Research, Applications. pp. 453–489.

Mazzarino, B., Mancini, M., 2009. The need for impulsivity & smoothness - improving HCI by qualitatively measuring new high-level human motion features. SIGMAP 2009 - International Conference on Signal Processing and Multimedia Applications, Proceedings. pp. 62–67.

Repp, B.H., 1995. Acoustics, perception, and production of legato articulation on a digital piano. J. Acoust. Soc. Am. 97 (6), 3862–3874.

Repp, B.H., 1997. Acoustics, perception, and production of legato articulation on a computer-controlled grand piano. J. Acoust. Soc. Am. 102 (3), 1878–1890.

Scherer, K.R., 1978. Personality inference from voice quality: the loud voice of extroversion. Eur. J. Soc. Psychol. 8 (4), 467–487.

Spence, C., 2011. Crossmodal correspondences: a tutorial review. Atten. Percept. Psychophys. 73 (4), 971–995. https://doi.org/10.3758/s13414-010-0073-7.

Volpe, G., Camurri, A., 2011. A system for embodied social active listening to sound and music content. J. Comput. Cult. Herit. 4 (1), 2:1–2:23.